

「データの時代の心理学」について(指定討論)

大阪電気通信大学
情報通信工学部 情報工学科
小森政嗣



指定討論:自己紹介

- 大阪電気通信大学 情報通信工学部 情報工学科
- 小森政嗣(こもりまさし)
- Twitterの投稿時刻、センチメント分析
- 携帯デバイスを利用した農村での「つながり」調査
- 活動量計を使った気分障害の研究
- 表情の時系列変化の研究
- などなど



2

「データの時代」?

- 素朴な疑問
- 精神物理学や知能検査の誕生の例を挙げるまでもなく、心理学はその黎明期から「データの時代」であったはずだが..
- 無論、念頭にあるのは“ビッグデータ”
 - 莫大な量の、かつ構造化されていないデータの蓄積
- “ビッグデータ”的「データ」はこれまでのデータと質的に違うのか? →おそらく違う
- データのとり方
- データ自体の性質
- データの解釈

この三点から
考えてみたい

3

変わったことその①:データのとり方の変化

- ノイジーなヒトの行動をどう扱うか
 - 従来一般的だった方法
 - 研究者によって統制・操作された環境で実験・調査を実施する
 - 研究者が作成した質問紙調査に回答を求める
 - 環境を統制することでアーチファクトを除去する
 - 「データの時代」の心理学
 - 経験サンプリング(experience sampling)や、ライフログデータ・SNSデータなどから得られた生態学的妥当性が高いデータ
 - 電子的に記録される
 - 大量のデータからロバストな相関を見出す

4

変わったことその①:データのとり方の変化

- 「行動」を直接的に記録できるようになった
 - 従来は工夫をこらした質問紙法調査や質問紙式の心理テスト法を用いて「行動」を記述していた
 - センサ技術・クラウド技術の発展によって、行動そのものの計測・記録ができるようになってきた
 - 配信された動画の面白さ
(評価の星の数☆☆☆ vs. 視聴切までの時間)
 - 睡眠尺度
(レトロスペクティブな報告 vs. ウェアラブル睡眠計)
 - 政治信条
(アンケートの回答 vs. Webのクリック)
- 時間的パースペクティブの変化
 - 従来はある時点を切り取ったデータ(環境で説明できない部分は「内的要因」で説明してきた)
 - 「データの時代」では、場合によっては一生を記録できるようになるかもしれない(環境で説明できる割合が大幅に増える)

5

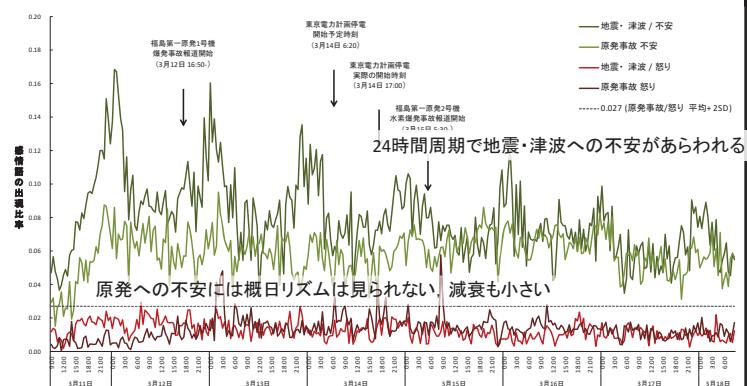
変わったことその②:データ自体の性質

- 「データの時代」に直面する統計的な難しさ
 - 時系列データ
 - 複雑な自己相関がある時系列データ
 - 時間的に隣り合うデータだけでなく離れたデータとも相関
 - 欠損値の多いデータ
 - 欠損値の割合が高い
 - なぜ欠損したのかがわからない
 - スパース性が高いデータ
 - 例: POSデータなど
- これまで心理学者が慣れ親しんでいた統計手法とは異なる手法が必要とされる

6

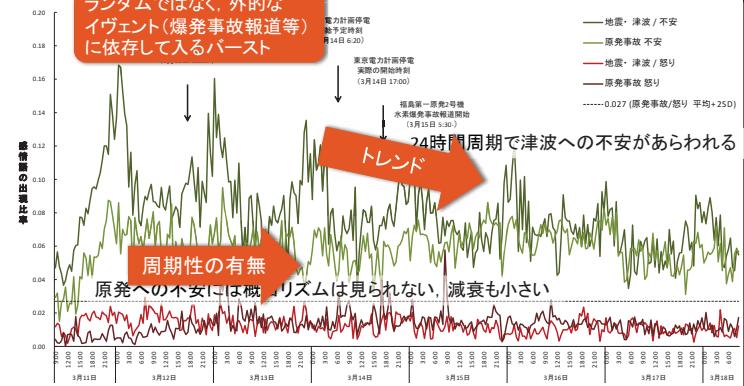
時系列データ解析の難しさ： 東日本大震災直後のツイッターのセンチメント分析

3.11後1週間の日本語全ツイートの分析



三浦・小森・松村・前田(2015). 東日本大震災時のネガティブ感情反応表出—大規模データによる検討—心理学研究, 86, 102-111.

時系列データ解析の難しさ： 東日本大震災直後のツイッターのセンチメント分析



三浦・小森・松村・前田(2015). 東日本大震災時のネガティブ感情反応表出—大規模データによる検討—心理学研究, 86, 102-111.

8

欠損値がありスパースなデータ

京丹後市常吉百貨店の来店データ

- ・地域の人の来店時間帯を30分毎に集計
- ・「共滞在パターン」を分析



9

10

変わったことその③：データの解釈

- ・「因果の理解」から「正確な予測」が主眼に？

少ないデータから
複雑なヒト・社会を
「因果関係」を積み
重ねることで理解
していく



大量のデータで見出
されたロバストな相
関を見出し、それを
用いて正確な「予測」
をする

「ビールとおむつ」の逸話
(理由は分からないがビールとおむつが一緒に売れる)

- ・ヒトや社会をブラックボックス化してしまうのではないかという懸念

実感として困ること

- ・「データのとり方」の大変さ
 - ・心理学者は大量のデータ処理のスキルがない
 - ・調査システムの作成も大変
- ・「データ自体の性質」の大変さ
 - ・特に時系列モデリングと欠損値の扱いの難しさ
- ・「データの解釈」の大変さ
 - ・妥当性の問題
 - ・既存の構成概念とどう接続すれば良いのか

発表者の方への質問その①： 心理学的な構成概念について

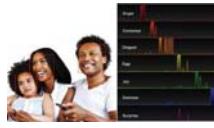
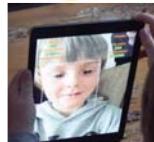
- ・心理学的な構成概念
- ・ここでは「外的な環境と独立して、行動に因果的影響を与える内的要因」のことを指すこととする
- ・欲求・情動・Big Five...etc.
- ・"ビッグデータ"から新たな構成概念が生まれるのか？
 - ・本来であれば、心理学の様々な構成概念を「流用」するだけでなく、"ビッグデータ"から新たな構成概念が誕生してほしい
 - ・そもそもヒトの行動を予測する上で、従来の心理学的な構成概念は(工学的に)意味があるのか？という疑問もある

心理学的な構成概念がうまく生かされている例

- 表情認識の分野では、基本6感情の考え方に基づく Facial Action Coding System (FACS) が用いられている

affectiva

EMOTIENT



Upper Face Action Units						
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7	
Brow Raise	Cheek Bow	Brow Lower	Upper Lid	Cheek Pull	Lid Tightener	
*AU 11	*AU 12	*AU 13	*AU 14	AU 15	AU 16	
Eye Brow Raiser	Eye Brow Lowerer	Eye Brow Outer Mover	Eye Brow Inner Mover	Eye Brow Dimpler		
*AU 41	*AU 42	*AU 43	*AU 44	AU 45	AU 46	
Lid Squeeze	Skin Tension	Eyes Squeezed	Septal Retractor	Blink	Wink	

Lower Face Action Units						
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14	
Nose Wrinkle	Upper Lip Dispenser	Nasolabial Fold	Lip Corner拉回	Cheek Dimpler		
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22	
—	—	—	—	—	—	
Lip Corner拉回	Lower Lip Dispenser	Chin Pressor	Lip Pressor	Lip Retractor		
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28	
—	—	—	—	—	—	
Lip Tightener	Lip Pressor	Lip Pucker	Mouth Stretch	Lip Suck		

FACS (Ekman, 1978)



13

発表者の方への質問その②: ヒトや社会が「ブラックボックス」になる心配

- 大量の変数間の相関で記述されたヒトや社会は、「ブラックボックス」に見えてしまう
- ブラックボックス化を回避するためには、人の心や社会をどのように記述すれば良いのか？
 - 内的要因と外界の相互作用をどのように記述していくべき良いのだろうか？
- 仮に、心理学的な構成概念を用いないが精度の良い予測をする「ブラックボックス」ができたとき、人々はそれでも心の仕組みを知りたいと思うのか？（弱気）